WE CLAIM:

1. A functional site descriptor that defines a spatial configuration for a functional site of a protein, which functional site corresponds to a biological function other than a divalent metal ion binding site, for application to an inexact, three dimensional structural model of a protein to determine whether the protein possesses the biological function corresponding to the functional site defined by the functional site descriptor, the functional site descriptor comprising a set of geometric constraints for one or more atoms in each of two or more amino acid residues comprising a functional site of a protein other than a divalent metal ion binding site, wherein at least one of said two or more amino acid residues is identified as a particular amino acid residue or set of amino acid residues, wherein said one or more atoms is selected from the group consisting of amide nitrogens, α-carbons, carbonyl carbons, and carbonyl oxygens within a polypeptide backbone, β-carbons of amino acid residues, and pseudoatoms, and wherein at least one of said one or more atoms is an amide nitrogen, an α-carbon, a β-carbon, or a carbonyl oxygen within a polypeptide backbone.

2. A functional site descriptor according to claim 1 wherein 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acid residues comprising the functional site are identified as particular amino acid residues or sets of amino acid residues.

3. A functional site descriptor according to claim 1 wherein the identity of an amino acid residue specified in the functional site descriptor is selected from the group consisting of Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, and Val.

4. A functional site descriptor according to claim 1 wherein the identity of an amino acid residue specified in the functional site descriptor comprises a set of two or more amino acid residue identities, wherein each of said amino acid residue identities is selected from the group consisting of Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, and Val.

5. A functional site descriptor according to claim 1 wherein each geometric constraint within the set of geometric constraints is selected from the group consisting of an atomic position specified by a set of three dimensional coordinates, an interatomic distance, and an interatomic bond angle.

6. A functional site descriptor according to claim 5 wherein at least one member of the set of geometric constraints is an atomic position specified by a set of three dimensional coordinates, wherein the atomic position can vary within a preselected RMSD.

7. A functional site descriptor according to claim 6 wherein the atomic position varies within an RMSD of less than about 3 Å.

8. A functional site descriptor according to claim 5 wherein at least one member of the set of geometric constraints is an interatomic distance range.

9. A functional site descriptor according to claim 5 wherein at least one member of the set of geometric constraints is an interatomic bond angle range.

10. A functional site descriptor according to claim 1 further comprising a conformational constraint.

11. A functional site descriptor according to claim 1 that comprises a set of geometric constraints with respect to at least one atom from each of 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acid residues that comprise the functional site corresponding to the functional site descriptor.

12. A functional site descriptor according to claim 1 wherein all of the atoms for which geometric constraints are provided comprise a part of the polypeptide backbone and are selected from the group consisting of α-carbons, amide nitrogens, carbonyl carbons, and carbonyl oxygens.

13. A functional site descriptor according to claim 1 wherein at least one of said one or more atoms is a pseudoatom.

14. A functional site descriptor according to claim 13 wherein the pseudoatom is a center of mass with respect to at least two atoms selected from the group consisting of atoms from one amino acid residue and atoms from at least two amino acid residues of the protein.

15. A functional site descriptor according to claim 1 implemented in electronic form.

16. A functional site descriptor according to claim 1 for a biological function selected from the group consisting of disulfide oxidoreductase activity, α/β hydrolase activity, phospholipase activity, and T1 ribonuclease activity.

17. A functional site descriptor according to claim 1 selected from the group consisting of a three atom functional site descriptor, a four atom functional site descriptor, a five atom functional site descriptor, a six atom functional site descriptor, a seven atom functional site descriptor, an eight atom functional site descriptor, a nine atom functional site descriptor, a ten atom functional site descriptor, an eleven atom functional site descriptor, a twelve atom functional site descriptor, a thirteen atom functional site descriptor, a fourteen atom functional site descriptor, and a fifteen atom functional site descriptor.

18. A functional site descriptor according to claim 1 wherein the functional site is selected from the group consisting of an active site of an enzyme, ligand binding domain, and a protein-protein interaction domain.

19. A functional site descriptor according to claim 18 wherein the ligand binding domain binds a ligand selected from the group consisting of a substrate, a co-factor, and an antigen.

20. A library of functional site descriptors, wherein the library comprises at least one functional site descriptor according to claim 1.

21. A library of functional site descriptors according to claim 20, wherein each of the functional site descriptors in the library is a functional site descriptor according to claim 1.

22. A library of functional site descriptors according to claim 20, wherein the library comprises at least two

functional site descriptors for at least one of the biological functions represented by the library.

23. Method of identifying a protein as having a particular biological function, the method comprising:

(a) applying a functional site descriptor according to claim 1 that correlates with the particular biological function to a structure of a protein; and

(b) identifying the protein as having the biological function if application of the functional site descriptor reveals that a portion of the structure of the protein matches the constraints of the functional site descriptor.

24. A method according to claim 23 wherein the structure of the protein is a high resolution structure.

25. A method according to claim 24 wherein the structure of the protein has been determined by x-ray crystallography or nuclear magnetic resonance.

26. A method according to claim 23 wherein the structure of the protein is a predicted structure.

27. A method according to claim 26 wherein the predicted structure is an inexact model of the structure of the protein.

28. A method according to claim 27 wherein the inexact model of the structure of the protein is produced by a computer running a computer program selected from the group consisting of an *ab initio* folding program, a threading program, and a homology modeling program.

29. A method according to claim 23 wherein the protein is an animal protein.

30. A method according to claim 29 wherein the animal protein is a mammalian protein.

31. A method according to claim 30 wherein the mammalian protein is a protein derived from a mammal selected from the group consisting of bovine, canine, equine, feline, ovine, and porcine animals.

32. A method according to claim 23 wherein the protein is a human protein.

33. A method according to claim 23 wherein the protein is a plant protein.

34. A method according to claim 23 wherein the protein is a prokaryotic protein.

35. A method according to claim 23 wherein the protein is a viral protein.

36. A method according to claim 23 wherein a plurality of functional site descriptors is applied to the structure of the protein.

37. A method according to claim 23 wherein the functional site descriptor is applied to a plurality of structures of the protein.

38. A method according to claim 23 wherein the functional site descriptor is applied to a structure of a plurality of proteins.

39. A method according to claim 23 wherein the functional site descriptor is applied to a plurality of structures for a plurality of proteins.

40. Method of identifying a protein as having a particular biological function, the method comprising:

(a) applying a functional site descriptor that correlates with the particular biological function to a predicted structure of the protein, wherein the functional site descriptor comprises a set of geometric constraints for one or more atoms in each of two or more amino acid residues comprising a functional site of a protein, wherein at least one of said two or more amino acid residues is identified as a particular amino acid residue or set of amino acid residues; and

(b) identifying the protein as having the biological function if application of the functional site descriptor reveals that a portion of the structure of the protein matches the constraints of the functional site descriptor.

41. A method according to claim 40 wherein the predicted structure is an inexact model of the structure of the protein.

42. A method according to claim 41 wherein the inexact model of the structure of the protein is produced by a computer running a computer program selected from the group consisting of an *ab initio* folding program, a threading program, and a homology modeling program.

43. Method of making a functional site descriptor that defines a spatial configuration for a functional site of a protein, which functional site corresponds to a biological

function other than a divalent metal ion binding site, for application to an inexact, three dimensional structural model of a protein to determine whether the protein possesses the biological function corresponding to the functional site defined by the functional site descriptor, the method comprising developing a set of geometric constraints for one or more atoms in each of two or more amino acid residues comprising a functional site of a protein other than a divalent metal ion binding site, wherein at least one of said two or more amino acid residues is identified as a particular amino acid residue or set of amino acid residues, wherein said one or more atoms is selected from the group consisting of amide nitrogens, α-carbons, carbonyl carbons, and carbonyl oxygens within a polypeptide backbone, β-carbons of amino acid residues, and pseudoatoms, and wherein at least one of said one or more atoms is an amide nitrogen, an α-carbon, a β-carbon, or a carbonyl oxygen within a polypeptide backbone.

44. A method according to claim 43 wherein the functional site is selected from the group consisting of an active site of an enzyme, a ligand binding domain, and a protein-protein interaction site.

45. A computer program product comprising a computer useable medium having computer program logic recorded thereon for creating a functional site descriptor for use in predicting a biological function of a protein, said computer program logic comprising computer program code logic configured to perform the operations of:

determining a set of geometric constraints for a functional site associated with a biological function of a protein;

modifying one or more geometric constraints of said set of geometric constraints to produce a modified set of geometric constraints;

comparing said modified set of geometric constraints to a data set of functional sites correlated with said biological function to determine whether said modified set of geometric constraint compares favorably with said data set of functional sites correlated with said biological function and, if so;

comparing said modified set of geometric constraint(s) to a data set of functional sites not correlated with said biological function to determine whether said modified set of geometric constraints compares favorably with said data set of functional sites not correlated with said biological function and, if so;

repeating said modifying and comparing operations to modify one or more of said geometric constraints of said set of geometric constraints to an extent that said modified set of geometric constraints compares favorably with said data set of functional sites correlated with said biological function without encompassing a predetermined amount of data sets not correlated with said biological function.

46. A computer program product according to claim 45, wherein said operation of determining a set of geometric constraints of a functional site correlated with a biological function of a protein comprises receiving said set of geometric constraints from at least one of the group of a data set of predetermined geometric constraints or from user input.

47. A computer program product according to claim 45, wherein said set of geometric constraints concerns one or more atoms in each of two or more amino acid residues comprising a functional site of a protein, wherein at least one of said two or more amino acid residues is identified as a particular

amino acid residue or set of amino acid residues, wherein said one or more atoms is selected from the group consisting of amide nitrogens, α-carbons, carbonyl carbons, and carbonyl oxygens within a polypeptide backbone, β-carbons of amino acid residues, and pseudoatoms, and wherein at least one of said one or more atoms is an amide nitrogen, an α-carbon, a β-carbon, or a carbonyl oxygen within a polypeptide backbone.

48. A computer program product according to claim 47 wherein said set of geometric constraints further comprises one or more geometric constraints with respect to one or more atoms or pseudoatoms of one or more amino acid residues that are adjacent to an amino acid residue of said two or more amino acid residues.

49. A computer program product according to claim 47, wherein said set of geometric constraints comprises geometric constraints selected from the group consisting of atomic positions specified by sets of three dimensional coordinates, interatomic distances, and interatomic bond angles.

50. A computer program product according to claim 47, wherein at least one of the geometric constraints of said set of geometric constraints comprises interatomic distances between one or more atoms and/or pseudoatoms of the amino acid residues of the functional site descriptor.

51. A computer program product according to claim 45, wherein said operation of modifying one or more geometric constraints of said set of geometric constraints to produce a modified set of geometric constraints comprises associating a predetermined variance with one or more of the geometric constraints.

52. A computer program product according to claim 45, wherein said operation of modifying one or more geometric constraints of said set of geometric constraints to produce a modified set of geometric constraints comprises:

computing an average value for a geometric constraint within the set of geometric constraints by determining values for said geometric constraint from two different proteins having functional sites that correlate with said biological function, and calculating said average value;

computing a standard deviation with respect to such geometric constraint; and

applying a multiplier to said computed standard deviation to generate said modified geometry.

ADD FIG